Advanced Methods for Passive Acoustic Detection, Classification, and Localization of Marine Mammals

Jonathan Klay
NOAA Pacific Marine Environmental Laboratory (PMEL)
2115 SE OSU Dr.
Newport, OR 97365
phone: (541) 867-0277 fax: (541) 867-3907 email: Jonathan.Klay@noaa.gov

David K. Mellinger
Cooperative Institute for Marine Resources Studies
Oregon State University (OSU)
2030 SE Marine Science Dr.
Newport, OR 97365
phone: (541) 867-0372 fax: (541) 867-3907 email: David.Mellinger@oregonstate.edu

Award Numbers: N0001413IP20051 / N0001412IP20052 / N0001411IP20086
web: ftp.pmel.noaa.gov with username ADCL

David J. Moretti
Naval Undersea Warfare Center (NUWC) Division, Newport
Attn: David Moretti, Code 70T
Bldg. 1351, 2nd floor
1176 Howell St
Newport RI 02842
phone: (401) 832-5749 fax: (401) 832-4441 email: David.Moretti@navy.mil

Award Numbers: N0001413WX20070 / N0001412WX20964 / N0001411WX21394

Steve W. Martin
SPAWAR Systems Center Pacific
53366 Front St.
San Diego, CA 92152-6551
phone: (619) 553-9882 email: Steve.W.Martin@navy.mil

Award Numbers: N0001413WX20071 / N0001412WX20033 / N0001411WX21401

Marie A. Roch
Department of Computer Science
San Diego State University (SDSU)
5500 Campanile Dr.
San Diego, CA 92182-7720
phone: (619) 594-5830 fax: (619) 594-6746 email: Marie.Roch@sdsu.edu

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **30 SEP 2013**	2. REPORT TYPE	3. DATES COVERED **00-00-2013 to 00-00-2013**
4. TITLE AND SUBTITLE **Advanced Methods for Passive Acoustic Detection, Classification, and Localization of Marine Mammals**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **NOAA Pacific Marine Environmental Laboratory (PMEL),2115 SE OSU Dr,Newport,OR,97365**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**21**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

LONG-TERM GOALS

For effective long-term passive acoustic monitoring of today's large data sets, automated algorithms must provide the ability to detect and classify marine mammal vocalizations and ultimately, in some cases, provide data for estimating the population density of the species present. In recent years, researchers have developed a number of algorithms for detecting calls and classifying them to species or species group (such as beaked whales). Algorithms must be robust in real ocean environments where non-Gaussian and non-stationary noise sources, especially vocalizations from similar species, pose significant challenges. In this project, we are developing improved methods for detection, classification, and localization of many types of marine mammal sounds.

OBJECTIVES

We are developing advanced real-time passive acoustic marine mammal detection, classification, and localization methods using a two-pronged approach: developing improved DCL algorithms, and developing standardized interfaces and software.

First, we are developing, testing, and characterizing advanced DCL algorithms for the following:

1. Echolocation click classification. Algorithms are being developed and tested for several species of beaked whales and small odontocetes.
2. Tonal signal detection and classification. Algorithms are being tested for several species of mysticetes and for small odontocetes.
3. Multi-sensor localization. Algorithms will be developed and tested on datasets containing sounds of both odontocetes and mysticetes.

Second, improved DCL software will be developed and both existing and new methods will be made available to users. The key contribution will be the development of four well-specified interfaces for detection, feature extraction, classification, and localization. We will implement the “front end” of these interfaces in widely-used and critical software packages, Ishmael and M3R, to supply acoustic data and metadata across the interfaces. Our “back end” implementations will encode DCL algorithms that can be plugged into any of the front ends to analyze acoustic data supplied across the interfaces. The aim is to make it simple for users to take advantage of these algorithms, and for developers to implement new methods in a simple, straightforward way and thus make them available to end users. We will conduct performance assessments of the improved algorithms and software interfaces using annotated data sets in the laboratory, and perform a demonstration using real time data at a US Navy instrumented range.

APPROACH

Odontocete click detection and classification

A multiclass support vector machine (SVM) classifier was previously developed (Jarvis et al. 2008). This classifier both detects and classifies echolocation clicks from five species of odontocetes, including Blainville's and Cuvier's beaked whales, Risso's dolphins, short-finned pilot whales, and sperm whales. Here Moretti's group, particularly S. Jarvis, will improve the SVM classifier by resolving confusion between species whose clicks overlap in frequency. The proposed work will investigate alternate feature sets to better separate species in the SVM's decision space.

The current real-time system of Roch et. al for odontocete click classification is based on Gaussian mixture models using cepstral feature vectors. Cepstral feature vectors provide a compact representation of the spectrum (Rabiner and Juang 1993) that let the system represent echolocation spectra using a reduced number of coefficients, providing a lower-dimensional feature space than using a standard representation of the spectrum. This system will be extended to cover more species and more recording/noise environments. In a separate project, Roch is working with personnel at Univ. Calif. San Diego on developing new features based on subspace models and improved noise compensation. The subspace models use hierarchical principal components analysis and random-projection trees (Freund et al. 2007) to learn new feature sets that will be used in place of cepstral feature vectors. The noise modeling will examine how to more effectively estimate background noise and compensate for it, taking into account interactions between noise and source (Ross 1976).

Tonal signal detection and classification

"Tonal signal" is a generic term for frequency-modulated calls such as baleen whale moans or odontocete whistles. Methods for detecting and classifying such sounds are being developed and applied to both odontocete whistles and baleen whale vocalizations, including minke (*Balaenoptera acutorostrata*), blue (*B. musculus*), and humpback (*Megaptera novaeangliae*) whales.

Odontocete clicks. The methods to be developed here determine the species associated with odontocete whistles that are extracted automatically via the *Silbido* tonal contour following system (Roch et al. 2010). Research led by Roch focuses on the areas of signal processing and *Silbido*'s search algorithm to further refine this algorithm. Echolocation clicks result in broad-band energy producing interfering peaks in the time-frequency domain. These will be mitigated by locating echolocation clicks through an existing detection algorithm (Soldevilla et al. 2008, Roch et al. 2011a) based on Teager energy (Kaiser 1990, Kandia and Stylianou 2006), and then removing it by interpolation.

In observing expert analysts classify whistles to species, we have noted that experts tend to comment on the general shape of a whistle. Extracted contours will be classified to species using hidden Markov models which are capable of modeling temporal transitions, thus exploiting the shape. HMMs have been used previously to classify signature whistles to groups, but a general approach requires more general models that can capture inter-specific variation. We propose segmenting whistles into components based upon easily identifiable landmarks (e.g. inflection points), and creating multiple models for components based upon cluster analysis.

Baleen whale vocalizations. Methods developed here for baleen whale detection and classification are based on automated detection and classification of minke whale boing vocalizations using tonal signal methods which have been previously applied to US Navy hydrophone data at PMRF (Mellinger et al. 2011; Martin et al. 2013). The minke boing call is complex, with multiple spectral components from very low frequencies to over 10 kHz. For hydrophones located in deep (>1 km) water such as PMRF, the dominant spectral component (DSC, described in Martin et al. 2013) is used for detection of the call as this component is typically the last component detected at long ranges (e.g. > 30 km). Minke boing call detection is used here has a first-stage detection step similar to the tonal detection processing described in Mellinger et al. 2011 in that a relatively narrow frequency band is used, with a requirement that a signal in this band exceed a threshold for a certain time period (e.g. 0.7 sec or more). Here, a second stage is also used which processes a slightly wider frequency band (1320 to 1450 Hz) to detect the onset frequency-modulation (FM) sweep component of the call. This is done to obtain a more accurate estimate of the start time of the call (compared to detection of the constant-frequency portion of the call) for later localization processing. The minke boing detection process

includes a third stage calculating the frequency with high spectral resolution (0.72 Hz per bin) of the DSC for each detected boing. The high-resolution DSC feature is also used in localization processing to help associate calls from individuals and in some cases to help track individuals over multiple hours.

Another approach being explored is to develop improved feature extraction methods that are based on processing units in the mammalian visual and auditory systems. It has been known for nearly 50 years that neurons in the visual cortex are sensitive to lines and surface edges in the visual field (Hubel and Wiesel 1962, Landy and Movshon 1991), and for at least 25 years that the auditory system has similar units for detecting frequency changes in tonal signals at specific frequencies (Mendelson and Cynader 1985). Mellinger and Martin will lead the effort to test some feature extraction and classification methods that use similar types of processing – specifically, developing processing units that respond to frequency change of a tonal signal within a narrow frequency range at specified FM rates, then modeling the time evolution of these units using a hidden Markov model (HMM) as described above.

Advanced localization algorithms. The first requirement for passive acoustic localization of marine mammals is the need to associate the detection of an individual signal as it is received across the array of widely spaced hydrophones. Moretti will lead the effort to develop a "nearest neighbor" approach to detection association. This approach will still use time difference of arrival (TDOA)/hyperbolic methods, but will not discard TDOA from pairs of detections when the normally requisite 3 detections are not achieved. Rather, detections from a given hydrophone will be associated with detections from all of its nearest neighbors and pair-wise TDOAs will be calculated.

Mellinger will also lead an effort to investigate an advanced localization method that employs the full cross-correlation function. The standard TDOA method extracts the position of the peak of the cross-correlation function between two hydrophones, and effectively ignores the rest of the cross-correlation. If the wrong peak is picked – which can happen easily due to multipath effects or, less commonly, interfering sounds – there is no information present to indicate that any other choice may have been nearly as good. Here we propose to use a system that uses the entire cross-correlation function for each hydrophone pair in finding the optimum location.

Software and interfaces. An Application Programming Interface (API) is a specification of a set of procedure calls (for objects, methods), data types (scalars, structures, classes, etc.), and protocols for use of the procedures and data types. A properly constructed and documented API makes it relatively simple for a developer to add new algorithms to an existing system. Systems with well-designed APIs permit users to add new functionality in a straightforward manner. Ishmael's (Mellinger 2001) interfaces for detection and localization comprise a relatively complex set of object class methods (procedure calls) and data types; although it is standardized, it is hardly straightforward or well-documented. The M3R system (Morrissey et al. 2006) has a format for standardized data serving and detection message passing using multicast over dedicated private networks. M3R also has a message-passing facility to share detection and classification results (i.e., notification of detection/classification events). Martin, Moretti, and Mellinger will lead the effort to develop and test APIs to make it relatively simple for developers to code new algorithms and test them in the Ishmael and M3R systems.

WORK COMPLETED

Meetings, data sharing site, and funding:

(1) We have had teleconference meetings approximately monthly to discuss both technical details and project logistics. We also had a face-to-face meeting during the International Workshop on Detection, Classification, Localization, and Density Estimation for Marine Mammals using Passive Acoustics in June 2013.

(2) We established a private Internet-accessible site for data sharing and have placed twelve data sets in it for use by project participants. This site is accessed at ftp.pmel.noaa.gov with username ADCL; contact the authors for required further login information. The site is private because some of the data, while not classified, is considered sensitive.

(3) Funding for the first two years of the project reached all project members. Year-3 funding has been slow to reach project members, particularly OSU and SDSU, because of delays in transferring funds from NOAA to OSU; essentially, about half of the funds didn't reach NOAA until late summer 2013. This delayed the ensuing transfer to SDSU, which is currently in process, with expected completion in October or November 2013.

Detection/Classification Algorithms

Advanced automated detection, classification and localization methods have been developed and applied to three species of baleen whales calls (humpback song, minke boing, and frequency downsweeps under 50 Hz typical of fin and sei whales). Improvements incorporated into baleen whale species call processing include a common processing front end (96 kHz sample rate, 16k FFT with Hanning window and 15k overlap). This processing front end has been coded for parallel processing utilizing multicore processors (also same as used for odontocete species detection and classification). Improvements to classification of minke whale boing calls this year include a wider frequency band (1320 to 1450 Hz), development of the high-resolution DSC frequency detector in this band, and a more accurate determination of each call's start time.

Detection of low frequency downsweeps (~35 to 20 Hz) for fin and sei whales has been implemented in the real-time processing string and is being utilized to find these species in available large data sets from PMRF. The processing follows the model utilized for the minke whale processing, with the same front end processing looking over a narrow frequency band for detections above the background estimate. The low frequency baleen species detector inclusion in the PMRF processing string aids in detecting and localizing fin/sei low-frequency downsweep type signals (e.g. 35 Hz to ~20 Hz). This allows rapid scanning for this type of low frequency baleen whale in data sets or in the real time system.

Humpback whale song unit processing has been initiated in the 200 Hz to 1200 Hz frequency band using the Generalized Power Law (GPL;Helble 2012). Humpback song unit automated localizations are being investigated via cross correlation of GPL outputs (vice spectrograms or time series) and are showing promise. The ability of the GPL processing to work in the presence of U.S. Navy MFAS activity has been initially reviewed with very good results.

Another new method for extracting features from frequency sweeps is under development. Termed "kernel-group spectrogram correlation," it uses a family of kernels operating on a normalized

spectrogram to find time-frequency locations at which tonal sounds, such as delphinid whistles or baleen whale tonal calls, are changing frequency (*df/dt*) at or near a specified rate.

Another component of our work on whistle classification has concentrated on increasing the purity of the automatically generated whistle clusters prior to training hidden Markov models.

A new method for employing support vector machines (SVM) to multi-class classification problems has also been developed. The new method is called the class-specific support vector machine (CS-SVM). A CS-SVM has been developed to classify click vocalizations from six species of odontocetes: Blainville's beaked whales (foraging clicks), Cuvier's beaked whales (foraging clicks), short-finned pilot whales, Risso's dolphins, sperm whales and pantropical spotted dolphins.

Localization Algorithms:
Time-difference-of-arrival (TDOA) estimation is typically done by cross-correlation. We developed a method for cross-correlation that uses clicks synthesized from times and amplitudes of peaks in the input signal.

Another localization method being developed uses intersecting hyperbolae from successive clicks in an echolocation sequence. Traditional multilateration requires the detection of a given signal by at least 3 widely separated sensors. Many odontocetes, including beaked whales, are known to emit vocalizations that are highly directive. While the source level of these vocalizations is estimated to be in excess of 200 dB re 1 μPA, the narrow beamwidth means that often only single sensors or pairs of sensors are ensonified at a time. Yet the animals are also know to sweep their heads as they forage for food and emit echolocation clicks. This head sweep over time allows for different nearby pairs of sensors to detect individual clicks. The TDOA between pairs of hydrophones defines a hyperbola in two dimensions (X-Y, with depth assumed constant). An algorithm was developed to plot the intersection of the hyperbolae from disparate pairs of hydrophones that receive different echolocation clicks in a sequence.

A model-based localization method has been developed and implemented, greatly improving localizing baleen species located further outside the PMRF hydrophone array. Minke whale localizations cannot be validated in the majority of cases. However, a minke localization has been validated with the limited visual sighting data available (field work with Tom Norris in 2009 resulted in one visual sighting in an area where PMRF range hydrophones localized a minke). Also, the use of mid-frequency active sonar detections has allowed validation of the localization processing methods by using ship GPS positional data with excellent results (acoustic locations typically within +/- 200 m of ship GPS positions).

Software:
The architecture for writing detection, classification, and localization modules has been completed and communication between Ishmael and a test module has been established. The architecture provides a translation library for each DCL platform supported that marshals data into a format that can be shared with other processes. Modules run as separate programs that share a limited region of memory with the DCL platform. This allows modules written on platforms that require separate processes (e.g. MATLAB, R) to be gracefully handled. User's designing classification modules will configure the DCL platform to send data to their module and make calls to a standard interface library. Results are sent back to the DCL platform in a similar manner.

RESULTS

Echolocation click detection

Our efforts in feature extraction for echolocation clicks have improved the robustness of species classification in conditions where the site or equipment differs from that seen in training data. To determine the impact of instrument and site variability, we selected two species that are relatively easy to classify due to distinct patterns in many of their echolocation clicks identified in previous work (Soldevilla et al. 2008). Risso's (*Grampus griseus*) and Pacific white-sided (*Lagenorhynchus obliquidens*) dolphins both have a series of spectral peaks and notches that enable excellent classification performance. Over 300,000 echolocation clicks were collected on autonomous high-frequency acoustic recording packages (Wiggins and Hildebrand, 2007) from six different sites throughout the Southern California Bight. Nine different series of preamplifiers were used across the deployments.

We began with a modification of our standard Monte Carlo method for evaluating classifier performance. Typically, we group echolocation click features from each acoustic encounter and ensure that they are all in the training or test data. A baseline mean error rate of 2.7%±2.5σ was obtained using our methods from Roch et al. (2011b). However, when the experiments were further constrained such that preamplifiers or sites were not split across the training and testing partitions (which also implies that acoustic encounters are not split), the error rate rose significantly to 20.9±18.1σ when partitioned by preamplifier and 25.9±28.1σ by site, increasing the error rate by nearly an order of magnitude. The preamplifier differences were unexpected given that the spectra were adjusted for a calibration of the preamplifier.

In previous work, we have found noise compensation methods to be ineffective, and hypothesized that weak echolocation clicks undetected by our signal processing chain may have been admitted to the noise estimation algorithm, thus contaminating the noise estimate. In this work, we established a weaker click detection threshold that was used for finding areas unlikely to contain echolocation clicks. Using the new noise estimates, the error rates fell significantly (Figure 1) with the preamplifier grouped error rate falling to 1.7±2.3σ and by site to 9.4±16.7σ. While more work remains to be done to diminish the effects of site variability, this work shows that noise compensation techniques can be extremely effective at diminishing the effects of instrument and site variability.

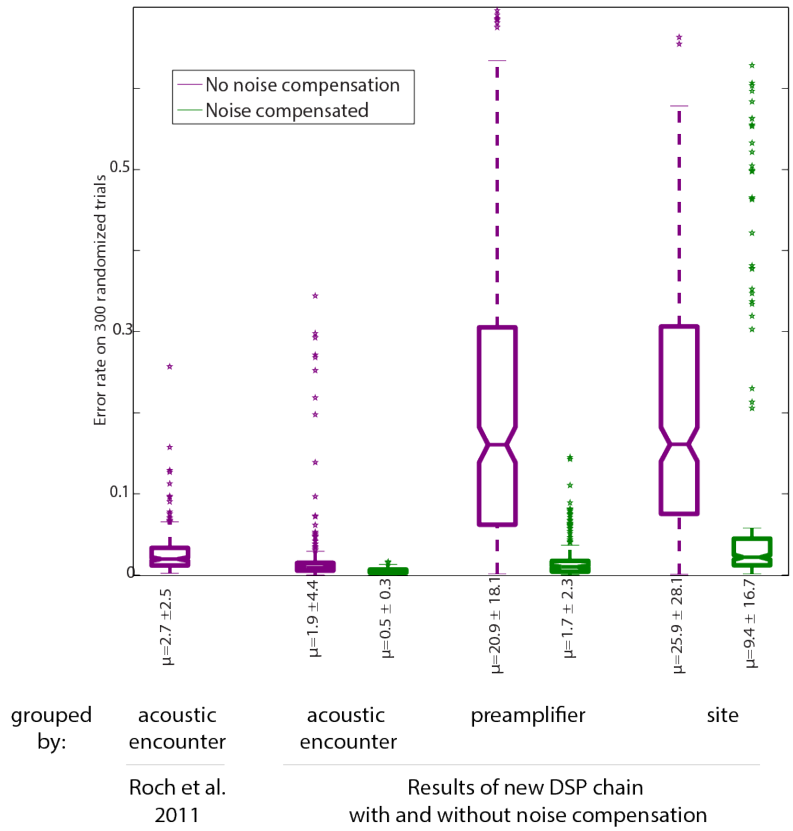


Figure 1. Error rates of 300 randomized echolocation click classification experiments comparing the baseline method and partitioning of the train/test data by encounter, preamplifier, and site with the new DSP chain, with and without noise compensation.

Odontocete whistle detection

Our work on whistle classification has focused on improving automated clustering of whistles. The root cause of tepid results from our hidden Markov model whistle classifiers was attributed to problems with the initial categories used to build the clusters and we focused our effort on improvements to clustering whistle components from individual species. To this end, we abandoned our current clustering method based on Deecke and Janik's ARTWARP (2006). While the dynamic time warping algorithm used by Deecke and Janik is a good approach for contour alignment, we hypothesized that modifications to the feature extraction would be beneficial. To that end, we developed a distortion function that better captured the differences between shapes of whistles. Traditional methods operate purely on frequency content. We computed derivatives of the contour and then normalized the features using a Z-score transform with the hypothesis that if shape is what really matters, two similar whistles at different frequencies should have low dissimilarity scores. A graph was constructed where each weighted edge represented the similarity between whistle components and the open-source package Gephi (Bastian et al. 2009) was used to organize the graph using a spring model and subsequently cluster the results (Figure 2). These new clusters represent an improvement over our previous clustering method and will be used to train the hidden Markov models in the coming year.

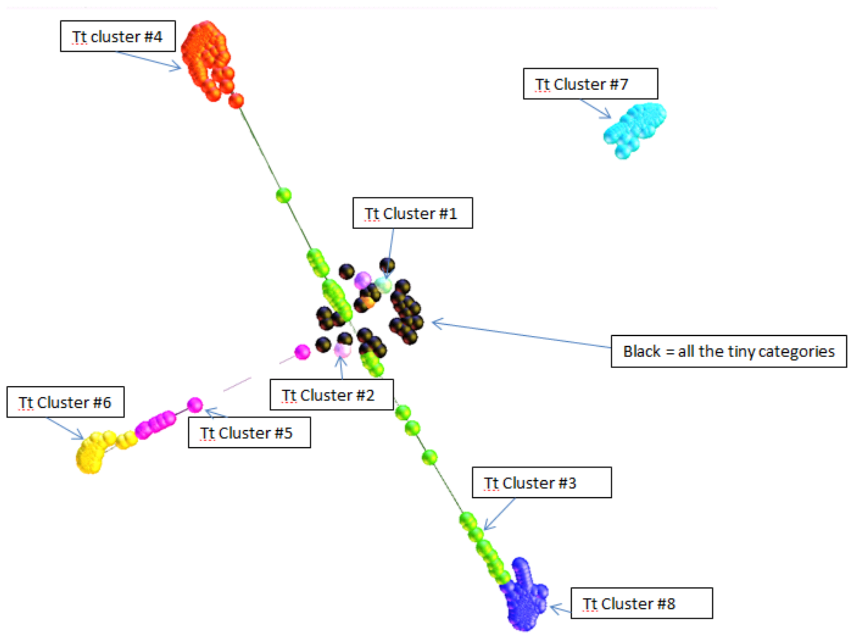


Figure 2 - Result of clustering bottlenose dolphin (Tursiops truncatus) components. Similarity metric derived from dynamic time warped comparison of Z-score normalized frequency contours and derivatives.

We also completed work on exploiting ridge information in spectrograms to help identify delphinid whistles (Kershenbaum and Roch, submitted). By looking at a spectrogram as a topological map, it is possible to examine the direction in gradient vectors and look for coherent regions where the signs of the gradient vectors swap. This algorithm has been incorporated into our whistle extraction algorithm *Silbido* (Roch et al. 2011a).

Another approach has been to use a new feature extraction method for whistles called highly-parallel multi-kernel correlation. First, whistles identified by the whistle detection stage (Mellinger et al. 2011) are normalized to remove background noise and to equalize differences in frequency response across different recording systems. An example of the spectrum of a common bottlenose dolphin whistle before and after normalization is shown in Fig. 3(a-b). Next, a series of kernels are generated for cross-correlation with the spectrogram of the whistle detected by Silbido; each kernel corresponds to a certain rate of frequency modulation (i.e., a certain slope) of the whistle at each instant. Fig. 3(c) shows one of these kernels and its corresponding feature map (Fig. 3(d)), which is the result of cross-correlating it with the normalized spectrogram. Since each kernel contains equal-strength positive and negative regions across each vertical time-slice in the spectrogram, it does not respond to echolocation clicks, as they do not produce a significant value in the correlation. (The maximum correlation result across the series of kernels gives the detected whistle, as in Fig. 3(e).) Features for use in classifiers are calculated by producing histograms of the kernel cross-correlation results and picking peaks in the kernels, with each peak corresponding to one rate of frequency modulation. These features represent the occurrence of certain FM rates; other features are appended to represent the frequency range of the whistle.

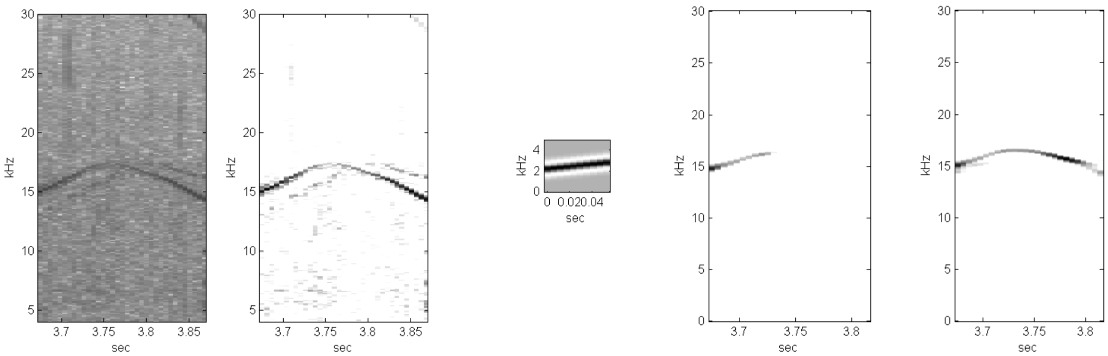


Figure 3. (a) Original and (b) normalized spectrogram of a common bottlenose dolphin whistle. (c) Spectrogram correlation kernel for the frequency modulation rate of 12 kHz/s. (d) Result of applying this one kernel to the normalized spectrogram. (e) Detected whistle after combining all kernel correlation results.

For testing and evaluating the proposed method, acoustic data from multiple surveys in the Southern California Bight will be used. The data for all species were recorded using towed and dipped hydrophone arrays and collected in the presence of single-species schools as determined by teams of experienced visual observers. The numbers of detected whistles for each species are listed in Table 1.

Table 1. Data used in whistle classification with highly-parallel multi-kernel correlation method.

Species	Number of files/locations	Number of detected whistles
Common bottlenose dolphin	3	372
Spinner dolphin	3	992
Melon-headed whale	3	354
Common dolphin	5	460

A preliminary implementation was made of the highly-parallel multi-kernel correlation method. In a classification using the detected whistles listed in Table 1, each species was modeled with a 16-mixture Gaussian mixture model (GMM). During testing, species were assumed to have a uniform prior distribution. The probability of a given whistle *i* occurring is given by

$$P\left(w_i \middle| M_{species}\right) = \sum_{m=1}^{16} b_m \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_m|^{\frac{1}{2}}} \times exp\left(-\frac{1}{2}(w_i - \mu_m)^T {\Sigma_m}^{-1} (w_i - \mu_m)\right)$$

where w_i represents the occurrence of whistle i, $M_{species}$ is the species-specific GMM, μ_m and Σ_m are species-specific mean and covariance matrices of the m^{th} normal distribution, b_m is the species-specific prior probability of the mixture, and d is the dimensionality of the feature space. The average error rate for 100 independent runs is 0.377. As shown, the highly-parallel multi-kernel correlation is able to extract whistle information.

CS-SVM classifier. Improvements have also been made to the class-specific support vector machine (CS-SVM) classifier presented last year. This classifier was developed to detect and classify the click vocalizations from six species of odonocetes. Alternate features and alternate feature estimation techniques have been investigated with the goal being to improve classification performance over that of the baseline CS-SVM.

To date, the CS-SVM has predominantly used click-level features. Single clicks are impulsive, short duration events and are relatively easy to detect and to extract features from. Candidate click-level features include time-domain features (envelope shape, peak time and amplitude), spectral domain feature (peak & notch frequencies), cepstral coefficients and wavelet features. However, most odontocete clicks are less than 0.5 ms long. There is only a limited amount of information available in such a short duration event. We investigated a number of alternate click-level features including envelope shape, peak and notch frequencies and cepstral coefficients. In the laboratory, with hand extracted features, these alternate feature sets resulted in classification performance that was very similar to that of our baseline feature set. This was not too surprising as the laboratory performance of the CS-SVM with baseline features is really quite good. The minimal change in performance with different click-level features was not sufficiently compelling to warrant modifying the baseline feature sets.

Click trains offer potential benefits over single clicks. They are a larger time-bandwidth product signal and, therefore, are theoretically more detectable. Beaked whales emit a nearly continuous stream of foraging clicks during their dives. The dives last several tens of minutes. The pattern of clicks versus time contains key information about the species. For example, at AUTEC, where three species of beaked whales have been observed, inter-click interval (ICI) is used by analysts to separate the species.

Automated click train feature extraction is more challenging because overlapping click trains, of same or different species are likely to be present. Extraction of inter-click interval (ICI), one of the baseline CS-SVM features, is a case in point. As noted above, ICI carries a lot of discriminating information but it can be deceptively hard to estimate automatically. The simplest approach to ICI estimation is to take the first difference between peak times but this quickly fails if multiple click trains are interleaved.

To gain the full benefit of using click trains, we must properly associate the stream of arriving clicks to the correct click trains automatically. This can be done using relative peak amplitudes to untangle click trains. The click associator selects clicks that are best matched in relative amplitude within a time window (Figure 4). The associated clicks can be used to form an improved estimate of a given feature or feature set, by averaging, for example. These better estimated input vectors are then input into the CS-SVM. The resulting improvement in classification performance in laboratory testing was significant. In a test, the baseline classifier had P_{corr}=0.76 for the *Zc* class, meaning that 76% of clicks were correctly associated. Using the input vectors for the CS-SVM generated by averaging the features observed over three associated clicks from the same click train, the classifier performance improved to P_{corr}=0.92 for the same *Zc* class. Again the classifier did not change; only the method for estimating the input features was improved.

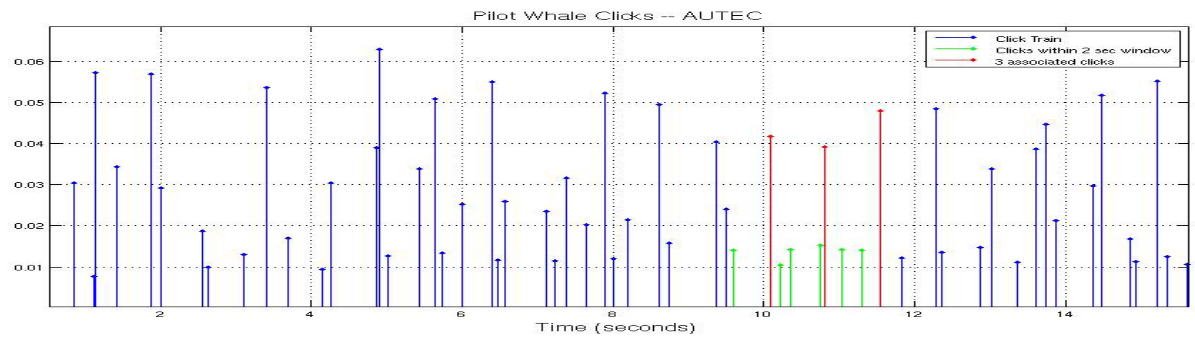


Figure 4. Associating clicks of interleaved click trains using relative amplitude.

Baleen whale call detection

Figure 5 shows the Islands of Kauai and Niihau, approximate locations of 41 of the 47 hydrophones used for minke boing localizations, and four localized minke whales indicated by yellow X symbols for 18 Feb 2013 at 0400 GMT. Each localized minke whale consists of multiple localizations over a user selectable temporal window, in this case one hour. Given the minke whales bimodal boing rate (one every ~0.5min or 5.5min) at the slower call rate there are 10 to 12 opportunities to localize an individual minke whale with ~10X as many if the whale is at the rapid boing rate. The model based localization is most accurate within the hydrophone array, however it is also showing good spatial grouping at distances as far as 20km from hydrophones in the east-west dimension. Localizations of different species of baleen whales are indicated with different symbols on the GUI (i.e. yellow “X” symbol for minke whales, orange “+” symbol for fin/sei/Bryde’s). The GUI also provides optional labels for the localizations which provide additional data for the species detections (e.g. detection time, number of hydrophones in the solution, the DSC frequency for minke species, and the least squares error of the localization).

A control window (not shown) for the GUI includes controls for other species of whales (currently beaked whales developed on other efforts, minke whales and <50Hz baleen whale calls) as well as controls for mid-frequency active sonar transmission localizations. The control window allows control for each localization category for how many hydrophones in the solution are required to display a localization (min 4 to max of 30), to display different species localizations or not, displaying the label for each localization, inclusion of a spatial cluster filter (to reject single spatially isolated localizations); and the temporal history to utilize for display of localization data. The time history can be scanned utilizing either a horizontal scroll bar located in the control window or by hitting the left/right arrows on the keyboard (the step size for each key kit is user variable from 1 s to 512 s doubling (or halving) with each keyboard + or – key entry).

Both the C++ detection and classification algorithm and C++ GUI display are working on both recorded data sets from PMRF (at 5X faster than real-time for 47 hydrophones) as well as operating on the M3R system (Morrissey et al. 2006) at PMRF in real time in February and August of 2013.

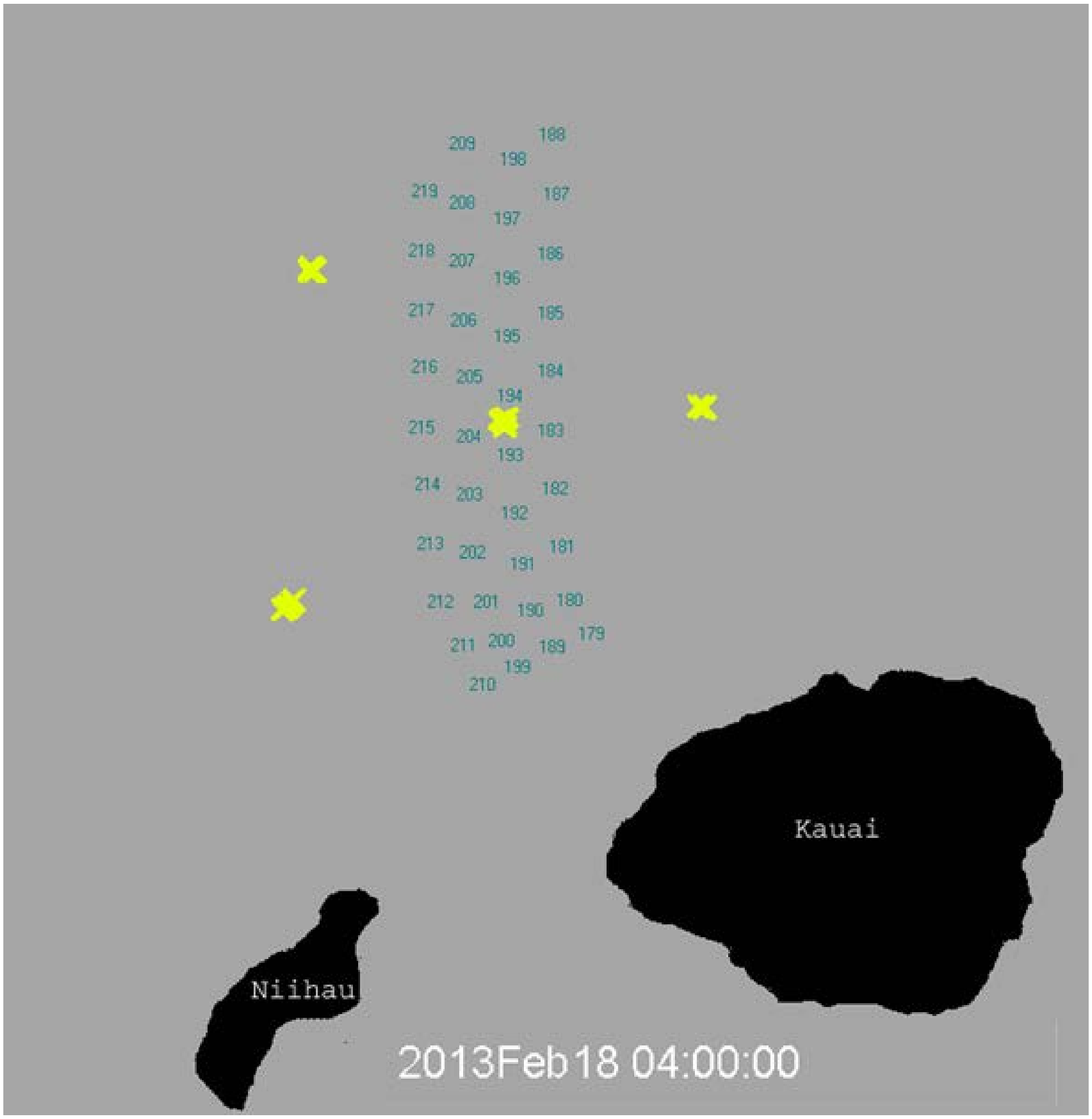


Figure 5. Graphical User Interface for visualization of automatic passive acoustic whale localizations showing approximate location of 41 range hydrophones and the relative locations of the Hawaiian Islands of Kauai and Niihau. This screen capture indicates a suspected four individual minke whales localized (multiple yellow X symbols per localization) over a 60 minute period between 0300 to 0400 GMT on 18 February 2013.

Humpback whale detection. The Generalized Power Law (GPL) detector (Nuttall 1996) has been successfully applied to humpback whale vocalizations (Helble et al. 2012). This processing has been applied to data from PMRF range hydrophones on this effort. The GPL processor is able to detect weak transient whale vocalizations in the presence of considerable anthropogenic and biological noise. This has proven to hold true even during periods of U.S. Navy mid-frequency active sonar transmissions typical in U.S. Navy training events. The current GPL detector is implemented in Matlab and operates approximately 60x faster than real-time for one hydrophone, and has robust thresholds for signal detection that do not need to be changed, even under drastically differing ocean-noise conditions. The ability of the GPL detector to work in this environment has been demonstrated using

PMRF data with promising preliminary localization results based upon cross correlation of GPL detections.

Beaked whale localization
For existing methods of localization using widely spaced, bottom-mounted sensors, typical of the Navy's undersea ranges, the localization of beaked whales presents a special problem. Two-dimensional localization of a marine mammal requires that its vocalization be received and detected on at least three hydrophones. Three-dimensional localization requires reception and detection on at least four hydrophones. However, the probability of receiving and detecting the echolocation clicks from beaked whales such as *Mesoplodon densirostris* (*Md*) simultaneously on more than one widely spaced hydrophone is low because their vocalizations are highly directional(Zimmer et al. 2008, Shaffer et al. 2013). Five algorithmic advances that significantly improve the ability to localize beaked whales have been made (Baggenstoss 2013).

(1) *Detection*: Development of a species-specific detector for detection of *Md* clicks at lower signal-to-noise ratio (SNR) than the existing hard-limited FFT detector currently employed [Morrissey, 2006, Jarvis, 2013].
(2) *Time-difference-of-arrival (TDOA) determination*: Development of a new means of eliminating spurious TDOA estimates which can arise in association of multiple overlapping click trains as received on a pair of widely spaced hydrophones
(3) *Time-difference-of-arrival (TDOA) tracking*: Development of a more reliable means of associating sequential TDOA measurements based on click matching. The method improves upon traditional TDOA tracking which creates tracks or trajectories from sets of TDOA measurements made at different times relying solely on the dynamics of the TDOA estimates.
(4) *TDOA association*. Development of a new means of associating two TDOA measurements made using different pairs of hydrophones, also based on click matching.
(5) *Localization*. Development of a means of accurate localization that incorporates the improved TDOA association and is able to resolve "disputes" that occur when a given TDOA estimate matches more than one positional solution.

Specifically, a detector based on replica correlation was shown to result in an increase in SNR (Figure 6). This is important due to the highly directional nature of beaked whale clicks. *Md*'s narrow beam pattern coupled with the geometry of the widely spaced hydrophone fields makes it unlike that more than one sensor will be approximately on-axis relative to the animal at any given time. The receive level of off-axis click can be reduced by 20+ dB relative to on-axis clicks. Any processing gain achieved improves the probability of click detection on the off-axis hydrophones.

A first order smoothed click map (SCM1) is then generated from the times of click detection (Figure 7). SCM1 from pairs of hydrophones are cross-correlated in the time domain to generate estimates of time-difference of arrival.

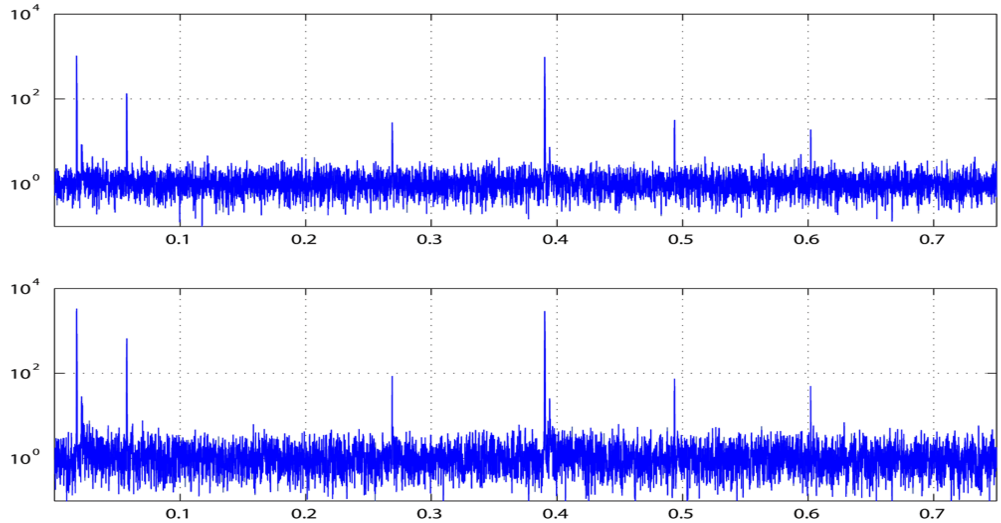


Figure 6: (Top) Instantaneous SNR estimate obtained prior to replica correlation. (Bottom) Same estimate of instantaneous SNR after replica correlation.

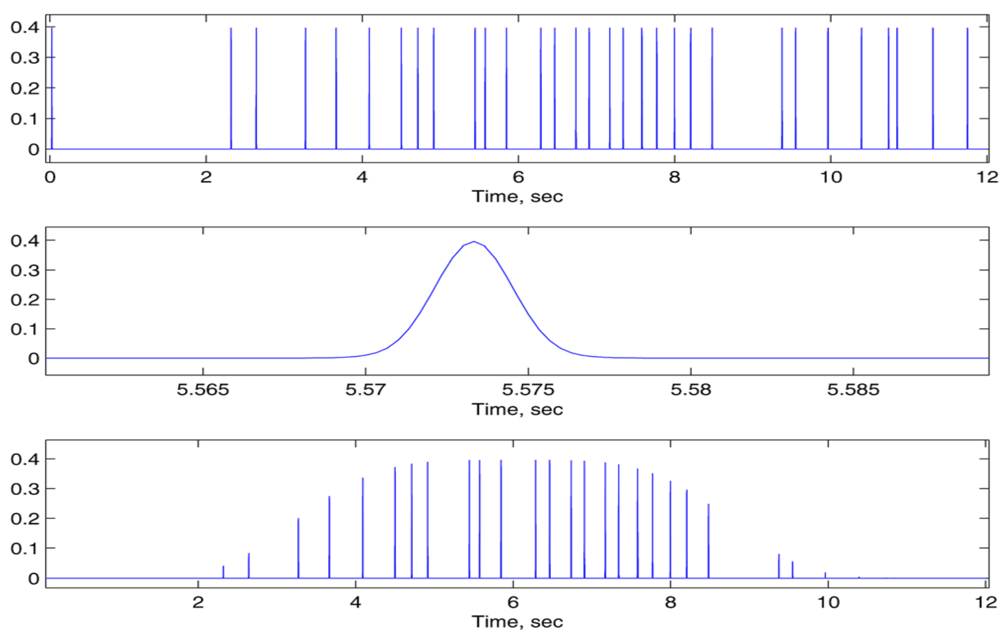


Figure 7. Example of click maps. (Top) First order smoothed click map (SCM1). (Middle) Zoom in on a single "pulse." (Bottom) Time-windowed SCM1.

The validity of an estimated TDOA is established by means of a second order clip map (SCM2), which is the shared click map exhibited by the pair of hydrophones assuming the estimated TDOA. Valid SCM2 (Figure 8-9) will exhibit the definitive characteristic of a Md click train, specifically an inter-click interval of approximately 0.37 seconds [Johnson, 2006].

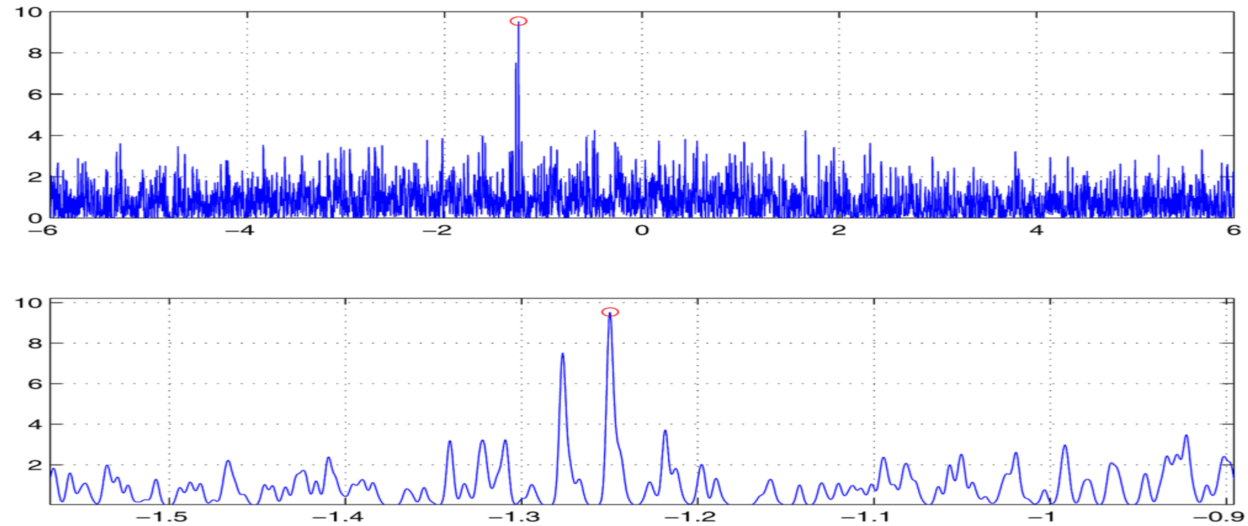


Figure 8: Example of correlation output between a pair of hydrophones. (Top) Full range and (Bottom) Zoom showing several peaks near -1.25 s. Red circle indicates largest peak. Which of these correlation peaks are valid?

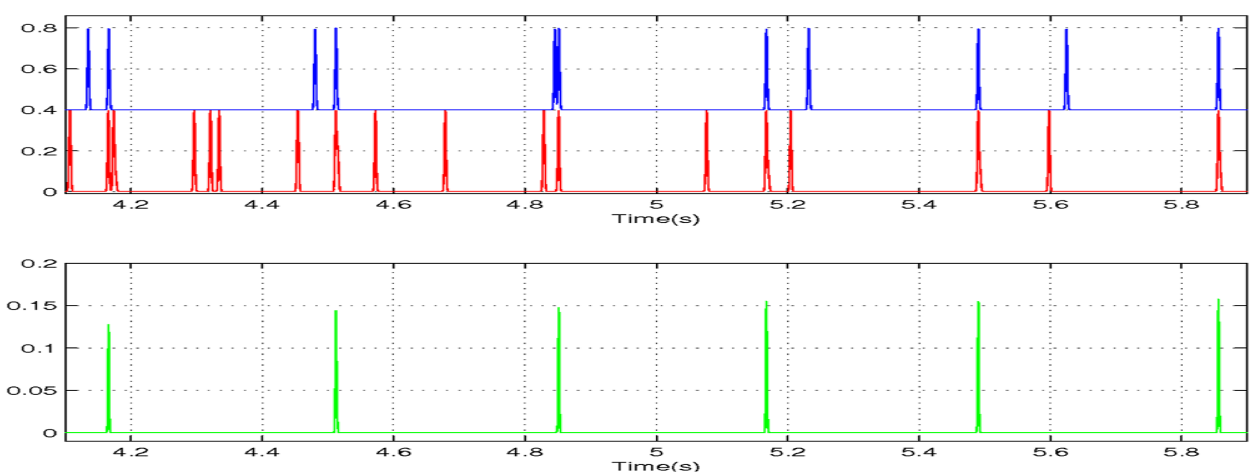
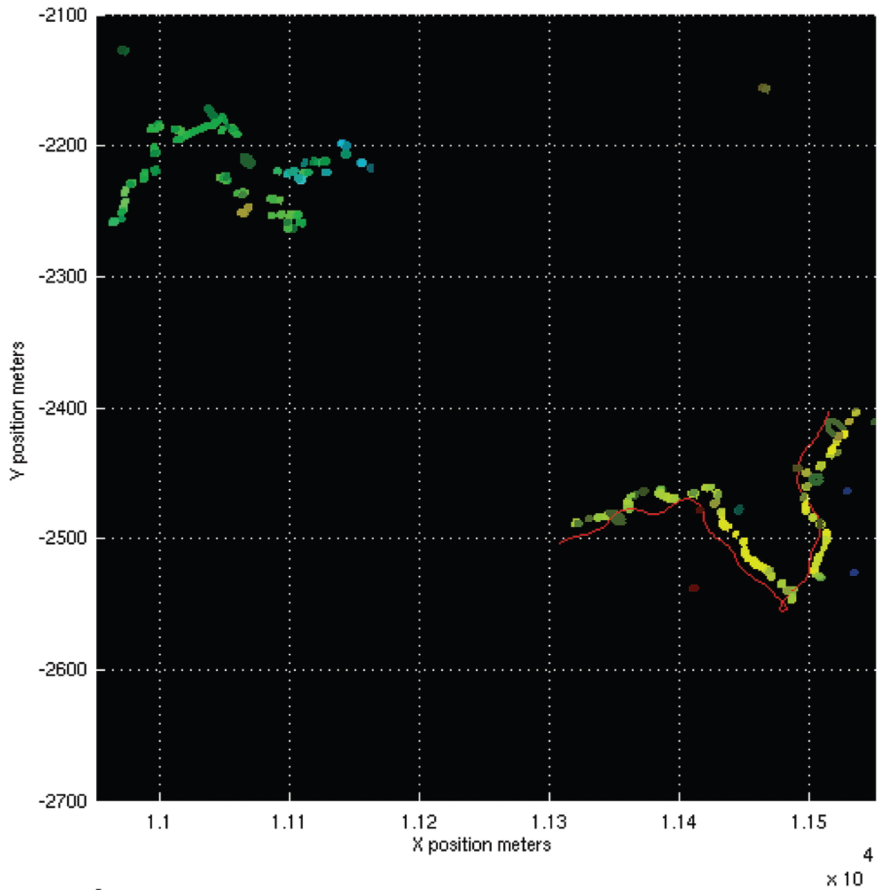


Figure 9: (Top) SCM1 for pair of hydrophones aligned at the time delay of the largest correlation peak from Figure 8. (Bottom) SCM2 derive by multiplying the SCM1 above. This click map displays a valid inter-click interval of approximately 0.33 seconds.

To perform localization at a particular point in time, the resultant TDOAs from pairs of hydrophone, which shared one phone, were scanned for TDOA estimates made within 4 s of the desired time. Then, all TDOA estimates that were located were associated by counting the number of clicks within a window that match between SCM2. This results in an inter-TDOA association measure. Localization was performed by forming an intensity surface I(x, y, z) for the hyperbola generated by the validated TDOA from associated pairs of hydrophones over a grid spaced 15 min in X and Y and 20m in depth. This was searched for local maxima. The local maxima detections were used as candidate solutions and iteratively refined (Figure 10).



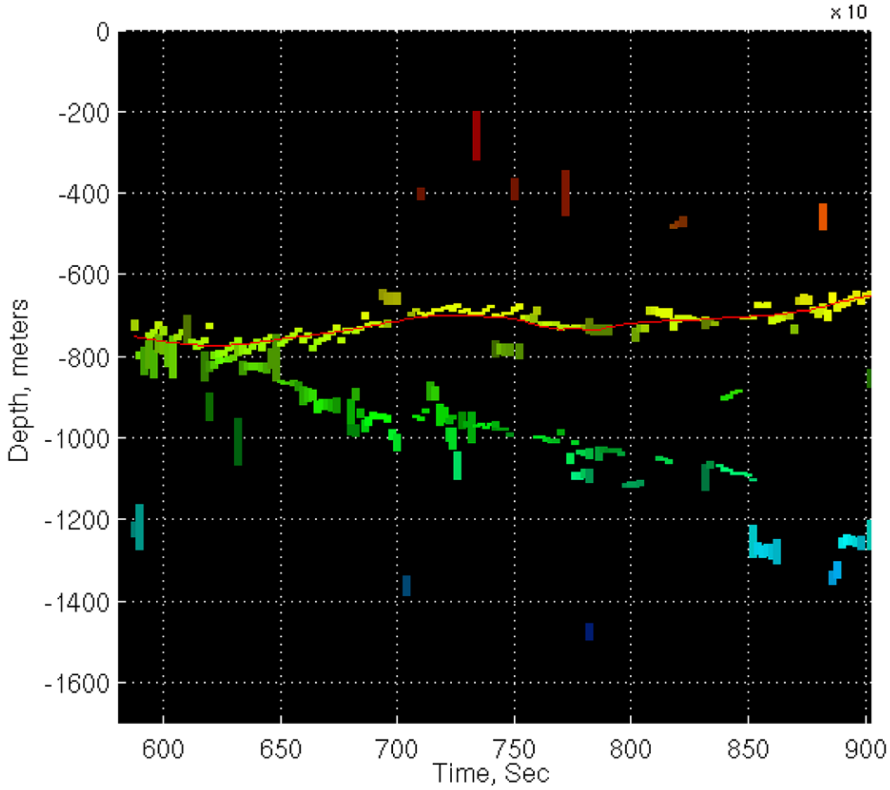


Figure 10: (Top) X-Y plot showing the localization of three separate Md. (Bottom) Depth tracks for these animals.

Baleen whale localization
Fin/sei type calls have been localized with good spatial grouping of calls within the hydrophone array. Additionally, on August 6, 2013 a low frequency baleen whale call was localized in real time (couple minute latency to localizations) crossing the range from east to west in the northern portion of the hydrophones. Figure 11 provides a zoom image of the GUI display showing approximate locations for three nearby hydrophones (184, 195 and 205) and 16 automatic DCL localizations where each localization had at least 13 hydrophones in the solution. The mean time between these localizations is 3.28 minutes (SD 0.71 min). The white line in Figure 2 approximates the distance traveled (10.32km) in the 46 minutes between the first and last localization plotted. This equates to an average swim speed of 13.3km/hr which while high is about one half of reported maximum swim speeds for Bryde's whale.

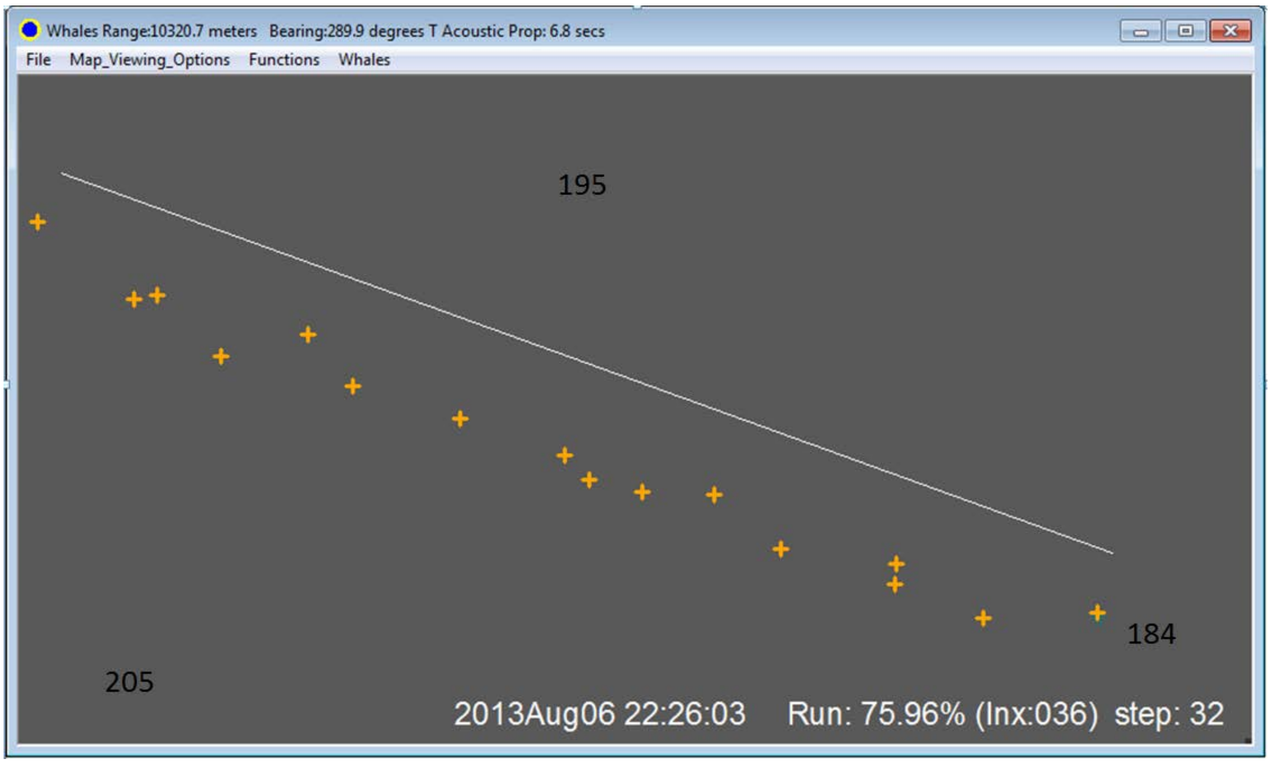


Figure 11 – Zoom of GUI display of a section of the PMRF offshore range (three hydrophone approximate locations shown) on Aug 6, 2013. Orange plus symbols are localizations of low frequency detections from a suspected Bryde's whale crossing the range from east (right) to west (left) from 21:30 to 22:16 GMT. White line is 10.24km in length.

Software
The Ishmael and PAMGUARD plug-in interface will allow people to write detection, classification, and localization algorithms to a single interface. Work has been completed on both PAMGUARD and Ishmael to permit bidirectional flow to user programs.

IMPACT/APPLICATIONS

For the Navy, passive acoustic monitoring (PAM) provides a means of long-term monitoring of many cetacean populations, especially over areas of high interest. Such areas are repeatedly subjected to

Navy exercises involving intense sounds, especially multi-ship mid-frequency active (MFA) sonar. Currently, required environmental monitoring is dependent primarily on visual line transect surveys that are costly and, in the case of aerial surveys, significantly dangerous. In both the areas critical to the Navy and in other areas critical to marine mammals, PAM is dependent on automated DCL methods. The advanced DCL algorithms being developed here will make PAM more effective and efficient; the algorithm implementations across standardized interfaces that handle both real-time and pre-recorded data streams from diverse platforms will make them available to Navy fleet operators as well as the wider marine mammal research community.

RELATED PROJECTS

"Passive Autonomous Acoustic Monitoring of Marine Mammals with Seagliders" (N00014-10-1-0387) award to Mellinger (and Klinck). The methods developed here are likely to be implemented in the Seaglider acoustic system for real-time detection and classification of marine mammal sounds.

"Acoustic Metadata Management and Transparent Access to Networked Oceanographic Data Sets" (NOPP N00014-11-1-0697) award to PI Marie Roch, Co-PI Simone Baumann-Pickering, John A. Hildebrand, et al. A metadata management system is being developed, which allows access to locally stored acoustic detections and metadata and links in a standardized way to external sources, such as oceanographic or ephemeris data. We will design our DCL plugins to provide outputs that can easily be stored in the acoustic metadata database.

RELATED PUBLICATIONS

Baggenstoss, P.M. 2013. Processing advances for localization of beaked whales using time difference of arrival. J. Acoust. Soc. Am. 133: 4065-4076.

Denes, S., J. Miksis-Olds, J. Nystuen, and D.K. Mellinger. In review. A comparison of marine mammal detections from two non-continuous autonomous acoustic recording systems. Submitted to J. Acoust. Soc. Am.

Jarvis, S.M., R.P. Morrissey, D.J. Moretti, J.A. Shaffer and N.A. DiMarzio. In review. Detection, localization and monitoring of marine mammals in open ocean environments using widely spaced bottom mounted hydrophones. Submitted to Mar. Tech. Soc. J.

Kershenbaum, A., M.A. Roch. In review. An image processing based paradigm for the extraction of tonal sounds in cetacean communications. Submitted to J. Acoust. Soc. Am; revision in review.

Lu, Y., Klinck, H., and Mellinger, D.K. In review. Noise reduction for better detection of beaked whale clicks. Submitted to J. Acoust. Soc. Am.

Martin, S.W., T.A. Marques, L. Thomas, R.P. Morrissey, S. Jarvis, N. DiMarzio, D. Moretti, and D.K. Mellinger. 2013. Estimating minke whale (Balaenoptera acutorostrata) boing sound density using passive acoustic sensors. Marine Mamm. Sci. 29:142-158, doi:10.1111/j.1748-7692.2011.00561.x.

Matsumoto, H., C. Jones, H. Klinck, D.K. Mellinger, and R.P. Dziak. 2013. Tracking beaked whales with a passive acoustic profiler float. J. Acoust. Soc. Am. 133:731-740.

Mellinger, D.K., M.A. Roch, E.-M. Nosal, and H. Klinck. In prep. Signal processing. Chapter for *Listening in the Ocean*, M. Lammers and W. Au, eds. To appear, late 2013 or early 2014.

REFERENCES

Bastian, M., Heymann, S. and Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. In Intl AAAI Conf Weblogs and Social Media. San José, CA.

Baggenstoss, P.M. (2013). Processing advances for localization of beaked whales using time difference of arrival. J. Acoust. Soc. Am. 133: 4065-4076.

Deecke, V. B. and Janik, V. M. (2006). Automated categorization of bioacoustic signals: avoiding perceptual pitfalls. J. Acous. Soc. Am. 199, 645-653.

Freund, Y., S. Dasgupta, M. Kabra, N. Verma. (2007) Learning the structure of manifolds using random projections, Adv. in Neural Info. Proc. Systems (NIPS), Vancouver, B.C., Canada.

Kershenbaum, A. and Roch, M. A. (submitted). An image processing based paradigm for the extraction of tonal sounds in cetacean communiations. J. Acous. Soc. Am.

Helble, T. A., G. R. Ierley, G. L. D'Spain, M. A. Roch, and J. A. Hildebrand (2012). "A generalized power-law detection algorithm for humpback whale vocalizations," J. Acoust. Soc. Am. 131(4), 2682-2699.

Hubel, D.H., and T.N. Wiesel. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *J. Physiol.* (London) 160:106-154.

Jarvis, S.M., N.A. DiMarzio, R.P. Morrissey, D.J. Moretti. (2008). A novel multi-class support vector machine classifier for automated classification of beaked whales and other small odontocetes. *Canad. Acoust.*, 36(1): 34-40.

Jarvis, S.M., (2012). A Novel Method For Multi-Class Classification Using Support Vector Machines. Ph.D. Dissertation. University of Massachusetts, Dartmouth.

Kaiser, J.F. (1990). On a simple algorithm to calculate the "energy" of a signal. Intl. Conf. Acoust., Speech, and Signal Process. IEEE, Albuquerque, NM: 381-384.

Kandia, V., Y. Stylianou. (2006). Detection of sperm whale clicks based on the Teager-Kaiser energy operator. *Appl. Acoust.*, 67(11-12): 1144-1163.

Landy, M.S., and J.A. Movshon, eds. (1991). *Computational Models of Visual Processing*. MIT Press, Cambridge.

Martin, A.Q., Marques, T.A., Thomas, L., Morrissey, R.P., Jarvis, S., DiMarzio, SN., and Moretti, D. (2013). Estimating minke whale (*Balaenoptera acutorostrata*) boing sound density using passive acoustic sensors. Marine Mammal Science 29(1): 142-158.

Mellinger, D.K. (2001). *Ishmael 1.0 User's Guide*. Natl. Oceanogr. Atmos. Admin. Tech. Memo. OAR–PMEL–120 (NOAA PMEL, Seattle). 30 pp.

Mellinger, D.K., Martin, S.W., Morrissey, R.P., Thomas, L. and Yosco, J. (2011). A method for detecting whistles, moans, and other frequency contour sounds. Jouranl of the Acoustical Society of America 129:4055-4061.

Mendelson, J.R., and M.S. Cynader. (1985) Sensitivity of cat auditory primary cortex (AI) neurons to the direction and rate of frequency modulation. *Brain Res.* 327: 331-335.

Morrissey, R.P., J. Ward, N. DiMarzio, S. Jarvis, and D.J. Moretti. (2006) Passive acoustic detection and localization of sperm whales (*Physeter macrocephalus*) in the Tongue of the Ocean. *Appl. Acoust*. 67: 1091-1105.

Norris R, Martin S, Thomas L, Yack T, Oswald JN, Nosal E-M, V Janik (2011) Acoustic Ecology and Behavior of Minke Whales in the Hawaiian and Marianas Islands: Localization, Abundance Estimation and Characterization of Minke Whale ”Boings” in eds Popper, Hawkins: The effects of noise on aquatic life: Second international congress, Springer.

Nuttall, A. (1996). “Near-optimum detection performance of power-law processors for random signals of unknown locations, structure, extent, and arbitrary strengths,” NUWC-NPT Technical Report, Newport, RI.

Oleson, E.M., Barlow, J., Gordon, J., Rankin, S., and Hildebrand, J.A. (2003). Low Frequency calls of Bryde’s whales. Marine Mammal Science 19(2):407-419.

Rabiner, L.R., B.H. Juang. (1986). An introduction to hidden Markov models. *IEEE ASSP Magazine*: 4-16.

Rabiner, L.R., and B.H. Juang. (1993). Fundamentals of Speech Recognition. Prentice Hall: Upper Saddle River, New Jersey.

Rankin, S. and Barlow, J. (2005). Source of the North Pacific ‘boing’ sound attributed to minke whales. Journal of the Acoustical Society of America 118: 3346-3351.

Roch, M.A., B. Patel, S. Rankin, Y. Barkley, M.S. Soldevilla, J.A. Hildebrand. (2010) Identifying delphinid whistle contours using graph search (A). *J. Acoust. Soc. Am.*, 127(3): 1936-1936.

Roch, M. A., Brandes, T. S., Patel, B., Barkley, Y., Baumann-Pickering, S. and Soldevilla, M. S. (2011a). Automated extraction of odontocete whistle contours. J. Acous. Soc. Am. 130, 2212-2223.

Roch, M. A., Klinck, H., Baumann-Pickering, S., Mellinger, D. K., Qui, S., Soldevilla, M. S. and Hildebrand, J. A. (2011b). Classification of echolocation clicks from odontocetes in the Southern California Bight. J. Acous. Soc. Am. 129, 467-475.

Ross, D. (1976) *Mechanics of Underwater Noise*. Pergamon: New York, xiv+375 pp.

Shaffer, J.W., D. Moretti, S. Jarvis, P. Tyack, and M. Johnson. (2013). Effective beam pattern of the Blainville's beaked whale (Mesoplodon densirostris) and implications for passive acoustic monitoring. J. Acoust. Soc. Am. 133:1770-1784.

Soldevilla, M.S. (2008) Risso's and Pacific white-sided dolphins in the Southern California Bight: Using echolocation clicks to study dolphin ecology. Ph.D. thesis, Univ. California, San Diego, La Jolla.

Soldevilla, M. S., Henderson, E. E., Campbell, G. S., Wiggins, S. M., Hildebrand, J. A. and Roch, M. A. (2008). Classification of Risso's and Pacific white-sided dolphins using spectral properties of echolocation clicks. J. Acous. Soc. Am. 124, 609-624.

Thompson, P.O. and Friedl, W.A. (1982). A long term study of low frequency sounds from several species of whales off Oahu, Hawaii. , Cetology 45:1-19.

Wiggins, S. M. and Hildebrand, J. A. (2007). High-frequency Acoustic Recording Package (HARP) for broad-band, long-term marine mammal monitoring. In Intl. Symp. Underwater Tech., pp. 551-557. Tokyo, Japan.

Zimmer, W.M.X., J. Harwood, P.L. Tyack, M.P. Johnson, and P.T. Madsen. (2008). Passive acoustic detection of deep-diving beaked whales. J. Acoust. Soc. Am. 124:2823-2832.